AGILENT TECHNOLOGIES, INC.
Legal Department, DL429
Intellectual Property Administration
P. O. Box 7599
Loveland, Colorado 80537-0599

ATTORNEY DOCKET NO. 10030215-1

AF

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s): James M. Minor

Serial No.: 10/821,829 **Examiner: Skowronek, Karlheinz R.**

Filing Date: April 9, 2004 **Group Art Unit: 1631**

Title: Methods and Systems for Evaluating and for Comaring Methods of Testing Tissue Samples

COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria VA 22313-1450

TRANSMITTAL OF APPEAL BRIEF

Sir:

Transmitted herewith is the Appeal Brief in this application with respect to the Notice of Appeal filed on

The fee for filing this Appeal Brief is (37 CFR 1.17(c)) **$500.00**.

(complete (a) or (b) as applicable)

The proceedings herein are for a patent application and the provisions of 37 CFR 1.136(a) apply.

☐ (a) Applicant petitions for an extension of time under 37 CFR 1.136 (fees: 37 CFR 1.17(a)(1)-(5)) for the total number of months checked below:

☐	one month	$ 120.00
☐	two months	$ 450.00
☐	three months	$1020.00
☐	four months	$1590.00

☐ The extension fee has already been filled in this application.

☒ (b) Applicant believes that no extension of term is required. However, this conditional petition is being made to provide for the possibility that applicant has inadvertently overlooked the need for a petition and fee for extension of time.

Please charge to Deposit Account **50-1078** the sum of $500.00. At any time during the pendency of this application, please charge any fees required or credit any overpayment to Deposit Account **50-1078** pursuant to 37 CFR 1.25.

A duplicate copy of this transmittal letter is enclosed.

☒ I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Date of Deposit: 6/21/07 **OR**

☐ I hereby certify that this paper is being facsimile transmitted to the Patent and Trademark Office on the date shown below.

Date of Facsimile:

Typed Name: Maria J. Sousa

Signature:

Respectfully submitted,

James M. Minor

By

Alan W. Cannon for John Brady
Attorney/Agent for Applicant(s)

Reg. No. 34,977

Date: 6/21/07

Telephone No. (408) 736-3554

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as first class mail in an envelope addressed to:
Mail Stop: Board of Patent Appeals and Interference, US Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450, on the date indicated below.
Name of Person Mailing: Maria Sousa

Signature: Maria Sousa
Date: 6/21/07

PATENT
Attorney Docket No.: 10030215-1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re application of:

James M. Minor

Application No.: 10/821,829

Filed: April 9, 2004

For: Methods and Systems for Evaluating and for Comparing Methods of Testing Tissue Samples

Assignee: Agilent Technologies, Inc.

Examiner: Skowronek, Karlheinz R.

Group Art Unit: 1631

Appeal Brief

Board of Patent Appeals and Interference
Alexandria, VA 22313-1450

APPEAL BRIEF (37 CFR §41.37)

This is an appeal from the Final Rejection in the Office Action mailed January 8, 2007, by the U.S. Patent and Trademark Office (USPTO) in the above referenced patent application. A Notice of Appeal was timely filed on May 8, 2007. Jurisdiction over this Appeal resides in the Board of Patent Appeals and Interferences (the Board) under 35 USC § 134.

Appellant/Applicant reserves the right to request an oral hearing.

I. Real Party in interest

Agilent Technologies, Inc., the assignee of the above referenced patent application is the real party in interest.

II. Related Appeals and Interferences

There are no related appeals or interferences.

III. Status of Claims

Claims 1-13 and 17-20 are pending in this application.

Claims 14-16 and 21-34 are canceled.

Claims 1-13 have been rejected under 35 USC § 102(b) as being anticipated by Singh et al., "Gene Expression Correlates of Clinical Prostrate Cancer Behavior", Cancer Cell, Vol. 1, March 2002.

Claims 1-10, 12-13 and 17-20 have been rejected under 35 USC § 102(e) as being anticipated by Crosby et al., U.S. Patent Publication No. 2003/0190689.

The rejections of Claims 1-13 and 17-20 are appealed.

IV. Status of Amendments

No amendments were filed after the Final Office Action dated January 8, 2007.

V. Summary of Claimed Subject Matter

A. Independent Claim 1 recites a method for rank ordering characteristic signatures of properties (stated in paragraph [0006] on page 3 of the specification). The method includes forming a plurality of characteristic signatures, which is described, for example at paragraph [0055], pages 16-17 of the specification. A characteristic signature may be a gene expression response signature (paragraph [0032}, page 8 of the specification), for example, of other profile of measurements of a particular characteristic of a target material (e.g., gas chromatograph readings, protein abundance analysis measurements, mass spectrometry readings, etc., see paragraph [0043}, page 11), wherein the measurements are taken from multiple locations (samples) along a pathway of the same target material containing tissue cells of interest and outlying tissues where it is relatively certain that none, or insignificant amounts of the cells of interest exist, see paragraph [0044], page 11 of the specification. A series of samples 108a, 108b, ..., 108n are taken along a line 106 which extends through the tissue cells of interest 104 and outlying tissues 102 on both sides, or any other trajectory where expected changes in density of the tissue cells of interest can be predicted or hypothesized, see paragraph [0044], page 11, of the specification, and Figs. 1-3.

For each tissue sample 108a, 108b, ..., 108n, a measurement of that tissue sample is taken for a particular property or characteristic that is to be characterized by the characteristic signature that is presently being formed, see paragraph [0049], page 14 of the specification. Thus for example, a first characteristic signature may be a vector of gene expression values for "gene A" measured for each of samples 108a – 108n, while a second characteristic signature may be generated for "gene B" as a vector of gene expression values for "gene B" , where a measurement of gene expression of gene B has been taken from each of samples 108a-108n, and so forth. Fig. 4 shows an idealized, schematic representation of a plot 400 for measured characteristic signatures (gene

response expression profiles, in this case) 404, 406 and 408, corresponding to three genes from the samples, wherein 404a is the gene expression value for the gene represented by characteristic signature 404, measured at sample a, 404b is the gene expression value for the gene represented by characteristic signature 404, measured at sample b, etc. (see also paragraph [0050], page 14 of the specification).

Independent Claim 1 further recites providing a trend profile of a second tissue measured property for the second type of tissue along the determined profile of locations through the tissue from which the samples were taken. For example, this may be a profile of a measure of disease activity across the sample tissue locations, or other characterization of the tissue of interest, e.g., see paragraph [0050], page 14 of the specification and Fig. 4, #402). Statistical analysis is performed with regard to the characteristic signatures relative to the trend profile, as those characteristic signatures that are closest to "synchronizing" or closely following the trend profile are considered to be related to, or involved in, the disease activity, or other characteristic of the tissue of interest that is measured by the trend profile, see paragraph [0051], page 15 of the specification, and Fig. 5. Rank ordering of the characteristic signatures, based on proximity to the trend profile is performed to identify those characteristic signatures that are considered to be most closely related to the trend profile, see paragraph [0055], page 17 of the specification and Fig. 5, #514.

B. Independent Claim 17 recites a computer readable medium carrying one or more sequences of instructions for rank ordering characteristic signatures of properties (stated in paragraph [0006] on page 3 of the specification), measured from a plurality of samples taken from a heterogeneous region, wherein a first portion of the heterogeneous tissue region has at least first and second types of tissue and is bordered by a second portion of the heterogeneous tissue region, wherein the second portion is considered to be devoid of the second type of tissue, and wherein the plurality of samples have been taken from successive locations along a determined profile of locations through the heterogeneous

tissue region, with at least one sample being taken from the second portion, wherein execution of one or more sequences of instructions by one or more processors causes the one or more processors to perform the method steps which have already been described in detail with regard to independent Claim 1 above.

C. Independent Claim 18 recites a system for rank ordering characteristic signatures of properties (stated in paragraph [0006] on page 3 of the specification) generated from tissue samples taken from a heterogeneous tissue region, wherein a first portion of the heterogeneous tissue region has at least first and second types of tissue and is bordered by a second portion of the heterogeneous tissue region, wherein the second portion is considered to be devoid of the second type of tissue. The system includes means for providing a trend profile of a second tissue measured property of the second type of tissue along a determined profile of locations through the heterogeneous tissue region from which tissues samples are taken as the sources of the characteristic signatures; means for performing statistical analysis on each of the plurality of characteristic signatures with regard to the provided trend profile; and means for rank ordering the plurality of characteristic signatures based on proximity to the trend profile as determined by the statistical analysis, each of which means may be embodiment by hardware described at paragraphs [0075] – [0076], page 25 of the specification, implementing multiple software modules for performing the method steps described in detail above with regard to Claim 1, see paragraph [0077], page 25 of the specification.

VI. Grounds of Rejection to be Reviewed on Appeal

A. Whether the invention as defined by Claims 1-13 is patentable under 35 USC § 102(b) over Singh et al., "Gene Expression Correlates of Clinical Prostrate Cancer Behavior", Cancer Cell, Vol. 1, March 2002 (hereafter, Singh et al.) because Singh et al. does not disclose or inherently possess each and every feature recited in Claims 1-13 of the instant application.

B. Whether the invention as defined by Claims 1-10, 12-13 and 17-20 is patentable under 35 USC § 102(b) over Crosby et al., U.S. Patent Publication No. 2003/0190689 (hereafter, Crosby et al.) because Crosby et al. does not disclose or inherently possess each and every feature recited in Claims 1-10, 12-13 and 17-20 of the instant application.

VII. Argument

A1. Rejection of Claims under 35 USC § 102(b) – Singh et al.

In the Final Office Action dated January 8, 2007 ("OA 01/08/07"), the Examiner rejected Claims 1-13 under 35 USC § 102(b) as being anticipated by Singh et al.

B1. Standard of Rejection under 35 USC § 102

Under 35 U.S.C. § 102, every limitation of a claim must identically appear in a single prior art reference for it to anticipate the claim. *In re Bond*, 910 F.2d 831,832, 15 U.S.P.Q.2D (BNA) 1566, 1567 (Fed. Cir. 1990).

C1(a). Application of Standard of Rejection under 35 USC § 102 to Claim 1

Appellant/Applicant submits that the Examiner has not met the burden of presenting a proper rejection under 35 U.S.C. § 102(b) because not every limitation of Claim 1 identically appears in the Singh et al. reference.

The Examiner asserted that Singh et al. discloses taking a plurality of samples from a heterogeneous tissue region and the formation of a plurality of characteristic signatures therefrom. The Examiner referred to page 209, column 1, second full paragraph of Singh et al. as support for these assertions, see page 3, second paragraph under the "Response to arguments" section of OA 01/08/07.

The Examiner further asserted that Singh et al. teaches the formation of a plurality of characteristic signatures. As support for this assertion, the Examiner referred to Figure 3 of Singh et al. indicating that this represents a plurality of characteristic signatures as a set of 12 genes, wherein the Examiner asserted that the measured expression of each gene listed in Figure 3 of Singh et al. forms a characteristic signature, and collectively, the genes constitute a plurality of characteristic signatures.

In the Advisory Action dated March 21, 2007 (AA 03/21/07), in response to Applicant's Request for Reconsideration filed February 26, 2007, the Examiner asserted that Singh et al. teaches taking a plurality of samples from tissue obtained from radical prostatectomy. The Examiner referred to page 208, column 1 of Singh et al., stating that this was interpreted to mean that normal tissue was bounded on opposite sides by tumorous tissue in the removed prostrate specimen and that this was interpreted to read on the claim recitation of a "plurality of samples taken from a heterogeneous region". The Examiner further indicated that Singh et al. teaches taking two samples, one of normal tissue and one of diseased tissue, from each of the 52 prostates obtained, and that this teaches the limitations of the claims reciting a plurality.

Appellant/Applicant respectfully traverses the Examiner's interpretations. First, Appellant/Applicant respectfully submits that page 208, column 1 of Singh et al. does not disclose use of 52 prostate specimens from a total of 65 prostate specimens, contrary to the Examiner's assertion. Rather, page 208, column 1 discloses 235 samples of prostate tumors and adjacent prostate tissue not containing tumor (normal), of which, 65 had cancer present on opposing sides of the OCT embedded specimens. Each of these samples was disclosed to have been reviewed.

Page 204, column 1 of Singh et al. discloses that high-quality expression profiles were successfully derived from 52 of the prostate tumors and 50 of the nontumor (normal) prostate samples. Thus, the "expression profiles" derived for these samples were derived

independently of one another, i.e., 52 expression profiles were derived for the 52 prostate tumor samples and 50 expression profiles were derived for the 50 normal samples. Each expression profile derived was derived for a single sample only. Thus, it is respectfully submitted that Singh et al. discloses the derivation of "gene expression signatures" or "gene expression profiles", and not "characteristic signatures" or "gene expression response signatures" or "gene expression response profiles" as recited in the present claims. Page 8, paragraphs [0031]-[0032] clearly distinguishes between these different types of signatures. Paragraph [0031] describes the type of expression profiles that are disclosed by Singh et al., which are the traditional "gene expression profiles", where a "gene expression profile" refers to gene expression values of a number of genes, typically from the same sample.

In contrast, a "characteristic signature", such as a gene expression response profile refers to a profile generated by expression values of the same gene over a number of samples. Claim 1 recites that each characteristic signature is formed of values for a particular property that has been measured from a plurality of samples taken from a heterogeneous tissue region.

Singh et al. clearly does not form a characteristic signature including a value from a normal tissue sample and a value from a tumor sample, contrary to the Examiner's assertion. Page 204, column 1 of Singh et al. discloses that "Genes were ranked according to their differential expression across the two classes (tumor versus normal), indicating that the expression profiles were each generated for a single sample (tumor or normal), not across a plurality of samples taken from the same tissue region.

Further, since the number of tumor samples (52) and nontumor samples (50) is unequal, this further supports that Singh et al. did not form characteristic signatures from values of property measured along different locations of a heterogeneous tissue region, since it would have been impossible to do so, in this case, for at least two of the tumor samples.

Instead, Singh et al. ranked gene expression values from each normal sample, relative to each other, and ranked gene expression values form each tumor sample, relative to each other.

Still further, neither locations of the tumor tissue samples nor the normal tissue samples of the Singh et al. study are identified or tracked so as to be able to correlate locations of gene expression values with locations of values along a trend profile of a second tissue measured property. Singh et al. does not provide a trend profile of a second tissue measure property for the second type of tissue along a determined profile of locations through a heterogeneous tissue region. That is, Singh et al. does not provide a trend profile of an activity level of the tumor tissue through both normal tissue and tumor tissue for any one particular sample, nor does Singh et al. provide a determined profile of any type of measure of the cancer tissue through the tissue samples measured. Rather, Singh et al. simply compares gene expression values for tumor tissues, relative to one another, to attempt to identify those genes that are significantly expressed within the tumor samples. Likewise, Singh et al. compares gene expression values for normal samples, relative to each other (normal samples) to attempt to find genes that are significantly expressed in normal tissues. This study is absolutely location independent, and is dependent only upon the types of tissues that are being examined. There is no disclosure or suggestion of comparing a profile including values of even one tumor sample and one normal sample from the same heterogeneous tissue sample, to a trend profile developed from values of another property at the same locations from which the samples were taken.

The Examiner further asserted that Singh et al. at page 206, column 1, paragraph 2, discloses that a readily detectable and statistically significant signature of GS (Gleason score) exists that shows a correlation between GS and the measured gene expression profiles. However, in the line following the referred to disclosure, Singh et al. discloses "The expression pattern of these genes separated tumors into distinct groups during

hierarchical clustering in both our initial and in a validation data set and grouped some of the intermediate grade tumors with high-grade tumors." The fact that Singh et al. refers to an "expression pattern" as composed of a plurality of "genes" further supports that Singh et al. generates "gene expression profiles", where each "gene expression profile" includes gene expression values of a number of different genes from the same sample, and that Singh et al. does not form a characteristic signature having values for the same gene across a number of locations measured in a single heterogeneous tissue region.

The Examiner further asserted in AA 03/21/07, that "Singh et al. also teach gene expression signature profile composed of a plurality of gene expression signatures of GS (p. 204, col. 2 para. 2, lines 1-5)." Upon referring to page 204, column 2, paragraph 2, lines 1-5 of Singh et al., Appellant/Applicant notes that this portion of the disclosure describes analyzing expression patterns within the 52 tumor samples. Accordingly, there were no gene expression values taken from the 50 normal samples for this portion of the study. Appellant/Applicant further respectfully submits that the "gene expression signature profiles" referred to by the Examiner and disclosed by Singh et al. as expression patterns are the traditional "gene expression profiles" described in the present application specification at page 8, paragraph [0031], as Singh et al. does not disclose or suggest forming a profile of gene expression values for one gene across a plurality of the samples. Still further, even if this were the case, and Singh et al. did form a profile of expression values for the same gene across a plurality of the 52 tumor samples, which Appellant/Applicant does not agree that Singh et al. discloses this, this would still not read on a characteristic signature formed of values for a particular property having been measure from a plurality of samples taken from a heterogeneous region, as the 52 tumor samples are not taken from a single heterogeneous region, but from 52 different patients. Still further, such a hypothetical characteristic signature (not disclosed by Singh et al.) would also not include a value from a normal sample, as Singh et al. clearly discloses that only the 52 tumor samples are being analyzed in this portion of the disclosure.

The Examiner further asserted in AA 03/21/07, that "Singh et al. provide a trend profile for the second type of tissue and perform a statistical analysis on each of the characteristic signatures." The Examiner asserted that Singh et al. teach a profile of a second tissue determined along the profile of locations in providing a "normal" gene expression profile, as disclosed at page 204, column 1. Appellant/Applicant does not understand this argument. Singh et al. discloses that genes were ranked according to their differential expression across the two classes (tumor versus normal). There is no disclosure of a trend profile along a determined profile of locations through a heterogeneous tissue region. Rather, Singh et al. merely ranks the expression values of genes within two different classes of tissues samples, i.e. a rank ordering of expression values within the normal samples group is performed and a rank ordering of expression values within the tumor samples group is performed. This has nothing whatsoever to do with correlating signatures with a trend profile based on locations of a heterogeneous tissue sample from which the values were taken.

The Examiner asserted in AA 03/21/07 that "A gene is interpreted to be a property and gene expression is interpreted to be characteristic signature." From this, the Examiner concluded that "Singh et al. teaches obtaining gene expression data (characteristic signatures) from normal and tumorous samples for 12,000 genes (properties). Appellant/Applicant respectfully submits that the Examiner's interpretation scheme referred to above falls apart logically in the portion of claim 1 that recites that a characteristic signature is formed of "values for a particular property having been measure from a plurality of samples taken from a heterogeneous tissue region". The 12,600 genes referred to by Singh et al. at page 204, column 1 are all genes from the same tissue sample as measured on a single microarray. Accordingly, each gene (property, according to the Examiner's scheme above) is associated with only one expression value, for only one tissue sample. Singh et al. does not disclose combining gene expression values across different microarrays, for a single gene, to form a characteristic signature. The Examiner's interpretation is faulty, because a "gene

expression profile" for 12,600 genes, thus having 12,600 gene expression values for 12,600 different genes, does not read on a characteristic signature having a plurality of values for a single gene having been measure from a plurality of samples taken from a single heterogeneous region, wherein the heterogeneous region contains diseased tissue in a first portion, and no diseased tissue in a second portion. Singh et al. does not form characteristic signatures as claimed. Also, Singh et al. analyzes the normal tissue samples separately from the tumor tissue samples.

The Examiner asserted in AA 03/21/07 that the Gleason score (GS) meets the recitation in Claim 1 of a "trend profile of a second tissue measured property for the second type of tissue along the determined profile of locations through the heterogeneous tissue region. The Examiner referred to Singh et al., page 204, column 2, paragraph 2, lines 2-5 as support for this assertion. This portion of the Singh et al. reference, as noted above, refers to analysis of "the expression patterns within the 52 tumors". Accordingly, it is not possible to consider a trend profile of a measured property for the second type of tissue (tumor) along the determined profile of locations through the heterogeneous tissue region, by this disclosure of Singh et al., because the determined profile of locations is defined in Claim 1 as including at least one sample being taken form the second portion which is devoid of tumor. Since Singh et al. analyzes only tumor tissues in the portion of the disclosure referred to by the Examiner, it follows that Singh et al. does not include a measurement from a nontumorous location in a trend profile.

The Examiner argued in AA 03/21/07 that Singh et al. takes samples from two successive locations in the same prostate specimen. Appellant/Applicant respectfully submits that the 52 tumor samples and the 50 normal samples are analyzed separately, since the gene expression values of the 52 tumor samples are rank ordered amongst themselves and the gene expression values of the 50 normal samples are rank ordered amongst themselves. Singh et al. fails to disclose forming a characteristic signature from a gene expression value for a gene (e.g., "gene A") from a tumor sample and a gene expression value for the

same gene (gene "A") from a normal sample adjacent the tumor sample, and then comparing such a characteristic signature with a trend profile containing values corresponding to the same two locations from which the gene expression values were read.

C1(b). Application of Standard of Rejection under 35 USC § 102 to Claim 6

Appellant/Applicant submits that the Examiner has not met the burden of presenting a proper rejection under 35 U.S.C. § 102(b) because not every limitation of claim 6 identically appears in the Singh et al. reference.

Claim 6 depends from claim 1 and, it is respectfully submitted, is allowable for at least the reasons provided above with regard to claim 1. Further, claim 6 recites in part "comparing each of the plurality of characteristic signatures with the provided trend profile by curve-fitting to a statistical regression function, wherein said curve-fitting determines the degree of proximity of each of the plurality of characteristic signatures to the provided trend profile."

In the Office Action dated 08/28/2006 (OA 08/28/2006), the Examiner referred to Singh et al., page 208, asserting that this discloses the above-recited claim language of Claim 6. However, the "K-nearest neighbor" techniques described by Singh et al. are used to determine Euclidean distances between gene expression values of individual genes. There is no disclosure of comparison to a trend profile, or of providing characteristic signatures as claimed.

D1. Conclusion

Based on the arguments set forth, Appellant/Applicant submits that under the Standard of Rejection under 35 USC § 102, Claims 1-13 are not properly anticipated by Singh et al.

A2. Rejection of Claims under 35 USC § 102(e) – Crosby et al.

In OA 01/08/07, the Examiner rejected Claims 1-10, 12-13 and 17-20 under 35 USC § 102(e) as being anticipated by Crosby et al.

B2. Standard of Rejection under 35 USC § 102

Under 35 U.S.C. § 102, every limitation of a claim must identically appear in a single prior art reference for it to anticipate the claim. *In re Bond*, 910 F.2d 831,832, 15 U.S.P.Q.2D (BNA) 1566, 1567 (Fed. Cir. 1990).

C2(a). Application of Standard of Rejection under 35 USC § 102 to Claim 1

Appellant/Applicant submits that the Examiner has not met the burden of presenting a proper rejection under 35 U.S.C. § 102(b) because not every limitation of Claim 1 identically appears in the Crosby et al. reference.

The Examiner in AA 03/21/07 asserted that Crosby et al. discloses at paragraph [0025] obtaining a plurality of samples form patients having positive and negative disease outcomes, which the Examiner considered to read on the claimed plurality of samples. Appellant/Applicant respectfully traverses this interpretation. Claim 1 specifically recites that the "plurality of samples" are taken "from a heterogeneous tissue region, wherein the heterogeneous tissue region includes a first portion having at least first and second types of tissue, bordered by a second portion, said second portion considered to be devoid of the second type of tissue. It is respectfully submitted that a plurality of samples taken from a single heterogeneous tissue region is distinct from the disclosure of Crosby et al. of taking samples from a plurality of patients, and thus a plurality of tissue regions.

The Examiner further asserted in AA 03/21/07 that Crosby et al. discloses measuring from a plurality of samples taken from a heterogeneous tissue region at paragraph [0080], where multiple sequential tissue slices are analyzed. Although Crosby et al. does disclose analyzing multiple sequential tissue slices, Crosby et al. also discloses at

paragraph [0080] that the tissue slices or cells are analyzed individually, either sequentially or in parallel. There is no disclosure of forming characteristic signatures as claimed. For example, there is no disclosure of forming a characteristic signature of protein activity values for a particular protein across multiple sequential slices or cells of a tissue region. Rather, particular cells having activated proteins are identified and then compared to a protein activity value of a normal cell. In contrast, the present invention forms a characteristic signature of values taken from both normal and diseased tissues at determined locations along a heterogeneous tissue region, and compares the characteristic signature to a trend profile of values from the same locations that the values for the characteristic signature were taken. This is neither disclosed nor suggested by Crosby et al.

The Examiner also referred to paragraph [0092] of Crosby et al. in AA 03/21/07, asserting that this discloses determining the correlation of protein activity (characteristic signature) and a disease outcome (trend profile). Again, the Examiner has incorrectly interpreted the recited elements "characteristic signature" and "trend profile". Paragraph [0092] indicates that cluster analysis may be performed on protein activity and a disease outcome, such as survival, or death. It is respectfully submitted that it would be nonsensical to consider forming a trend profile of values along determined locations of a heterogeneous tissue region if the values for the trend profile were either death or survival, since either the tissue survives or it dies. Accordingly, only a flat trend profile could be prepared according to the Examiner's interpretation, either having all values of "death" or all values of "survival". This would result in a flat trend profile, which would not be useful for analysis.

Additionally, as already noted, Crosby et al. does not disclose forming a characteristic signature of values for a single particular property, across samples at determined locations of a heterogeneous tissue.

C2(b). Application of Standard of Rejection under 35 USC § 102 to Claim 17

Appellant/Applicant submits that the Examiner has not met the burden of presenting a proper rejection under 35 U.S.C. § 102(b) because not every limitation of claim 17 identically appears in the Crosby et al. reference.

It is respectfully submitted that Claim 17 is allowable for at least the same reasons provided above with regard to Claim 1, as Claim 17 contains all of the claim limitations distinguished above with regard to Claim 1.

Further, the Examiner in AA 03/21/07 asserted that Crosby et al. discloses at paragraph [0013] automatic analysis using high-throughput automation. It is respectfully submitted that paragraph [0013] of Crosby et al. is a part of the background section that identifies what would be desirable, and does not disclose anything concrete, but rather indicates what would be desirable to achieve. Further, paragraph [0013] does not disclose a computer readable medium.

C2(c). Application of Standard of Rejection under 35 USC § 102 to Claim 18

Appellant/Applicant submits that the Examiner has not met the burden of presenting a proper rejection under 35 U.S.C. § 102(b) because not every limitation of claim 18 identically appears in the Crosby et al. reference.

It is respectfully submitted that Claim 18 is allowable for at least the same reasons provided above with regard to Claim 1, as Claim 18 contains claim limitations distinguished above with regard to Claim 1.

Further, the Examiner in AA 03/21/07 asserted that Crosby et al. discloses at paragraph [0013] automatic analysis using high-throughput automation. It is respectfully submitted that paragraph [0013] of Crosby et al. is a part of the background section that identifies what would be desirable, and does not disclose anything concrete, but rather indicates

what would be desirable to achieve. Further, paragraph [0013] does not disclose a system.

Still further, since Crosby et al. does not provide a trend profile as claimed, for reasons discussed above with regard to Claim 1, it follows that Crosby et al. does not provide means for providing a trend profile as part of a system.

Since Crosby et al. does not perform statistical analysis on a plurality of characteristic signatures with regard to a trend profile as claimed, it follows that Crosby et al. does not provide means for performing statistical analysis on each of the characteristic signatures with regard to the provided trend profile.

D2. Conclusion

Based on the arguments set forth, Appellant/Applicant submits that under the Standard of Rejection under 35 USC § 102, Claims 1-10, 12-13 and 17-20 are not properly anticipated by Crosby et al.

E. Conclusion

The Examiner has failed to show anticipation of Claims 1-13 under 35 USC § 102(b) over Singh et al.

The Examiner has failed to show anticipation of Claims 1-10, 12-13 and 17-20 under 35 USC § 102(e) over Crosby et al.

Accordingly, Appellant/Applicant requests that the Board of Patent Appeals and Interferences reverse the rejection of Claims 1-13 and 17-20 on the grounds set forth herein.

Respectfully submitted,

Date: 6/21/07

By: _____
Alan W. Cannon for John Brady
Registration No. 34,977

John Brady
Agilent Technologies, Inc.
Legal Department, DL429
Intellectual Property Administration
P.O. Box 7599
Loveland, CO 80537-0599
Telephone: (408) 553-3584
Facsimile: (408) 553-2365

VIII. Claims Appendix

This Appendix contains the Claims involved in the Appeal:

1. A method for rank ordering characteristic signatures of properties, said method comprising the steps of:

forming a plurality of characteristic signatures, each said characteristic signature being formed of values for a particular property having been measured from a plurality of samples taken from a heterogeneous tissue region, wherein the heterogeneous tissue region includes a first portion having at least first and second types of tissue, bordered by a second portion, said second portion considered to be devoid of the second type of tissue, wherein the plurality of samples have been taken from successive locations along a determined profile of locations through the heterogeneous tissue region, with at least one sample being taken from the second portion, and wherein each said characteristic signatures are formed from different ones of said particular properties, respectively;

providing a trend profile of a second tissue measured property for the second type of tissue along the determined profile of locations through the heterogeneous tissue region;

performing statistical analysis on each of the plurality of characteristic signatures with regard to the provided trend profile; and

rank ordering the plurality of characteristic signatures based on proximity to the trend profile as determined by the statistical analysis.

2. The method of claim 1, further comprising the step of:

measuring the specific property for each of the plurality of samples for at least one of characteristic signatures.

3. The method of claim 1, further comprising the steps of:

providing the heterogeneous tissue region: and

taking the plurality of samples from the heterogeneous tissue region.

4. The method of claim 3, further comprising the step of:

measuring each specific property for each of the plurality of samples for each respective characteristic signature.

5. The method of claim 1, wherein the step of forming a plurality of characteristic signatures includes normalizing each of the plurality of characteristic signatures with respect to a baseline reference signature, said baseline reference signature corresponding to a measured property of a sample taken from the second portion.

6. The method of claim 1, wherein the step of performing statistical analysis includes:

comparing each of the plurality of characteristic signatures with the provided trend profile by curve-fitting to a statistical regression function, wherein said curve-fitting determines the degree of proximity of each of the plurality of characteristic signatures to the provided trend profile.

7. The method of claim 1, wherein the step of performing statistical analysis includes:

calculating a p-value with regard to each of the plurality of characteristic signatures, to test the null hypothesis between each of the plurality of characteristic signatures and the provided trend profile.

8. The method of claim 1, wherein the step of performing statistical analysis is done in one-, two- or three-dimensional space.

9. The method of claim 1, wherein the first type of tissue is healthy tissue.

10. The method of claim 1, wherein the second type of tissue is diseased tissue.

11. The method of claim 1, wherein one of the specific properties is an expression level of a gene.

12. The method of claim 2, wherein the step of measuring the specific property comprises measuring different specific properties across the samples to form a plurality of characteristic signatures by processing each of the plurality of samples using a microarray technique.

13. The method of claim 13, wherein said processing comprises processing each of the plurality of samples on a single two-color microarray, two single-color microarrays or both.

17. A computer readable medium carrying one or more sequences of instructions for rank ordering characteristic signatures of properties measured from a plurality of samples taken from a heterogeneous region, wherein a first portion of the heterogeneous tissue region has at least first and second types of tissue and is bordered by a second portion of the heterogeneous tissue region, wherein the second portion is considered to be devoid of the second type of tissue, and wherein the plurality of samples have been taken from successive locations along a determined profile of locations through the heterogeneous tissue region, with at least one sample being taken from the second portion, wherein execution of one or more sequences of instructions by one or more processors causes the one or more processors to perform the steps of:

forming a plurality of characteristic signatures, each said characteristic signature being formed of values for a particular property having been measured from the plurality of samples;

providing a trend profile of a second tissue measured property of the second type of tissue along the determined profile of locations through the heterogeneous tissue region;

performing statistical analysis on each of the plurality of characteristic signatures with regard to the provided trend profile; and

rank ordering the plurality of characteristic signatures based on proximity to the trend profile as determined by the statistical analysis.

18. A system for rank ordering characteristic signatures of properties generated from tissue samples taken from a heterogeneous tissue region, wherein a first portion of the heterogeneous tissue region has at least first and second types of tissue and is bordered by a second portion of the heterogeneous tissue region, wherein the second portion is considered to be devoid of the second type of tissue, the system comprising:

means for providing a trend profile of a second tissue measured property of the second type of tissue along a determined profile of locations through the heterogeneous tissue region from which tissues samples are taken as the sources of the characteristic signatures;

means for performing statistical analysis on each of the plurality of characteristic signatures with regard to the provided trend profile; and

means for rank ordering the plurality of characteristic signatures based on proximity to the trend profile as determined by the statistical analysis.

19. The system of claim 18, further comprising

means for forming the plurality of characteristic signatures based on measurements of values for a particular property measured from the plurality of samples, each of said characteristic signatures formed from values for a particular property different from the particular properties measured to form others of the characteristic signatures.

20. The system of claim 18, further comprising:

means for measuring at least one said specific property for each of the plurality of samples.

IX. Evidence Appendix

This Appendix contains copies of:

A. Singh et al., "Gene expression correlates of clinical prostate cancer behavior", Cancer Cell: March 2002, Vol. 1, 203-209.
This reference was entered by Examiner Karlheinz R. Skowronek in the Office Action dated 08/28/2006.

B. Crosby et al., U.S. Patent Publication No. 2003/0190689 A1.
This reference was entered by Examiner Karlheinz R. Skowronek in the Office Action dated 08/28/2006.

X. Related Proceedings Appendix

This Appendix has no copies of decisions in related proceedings because there are no related proceedings to this Appeal.

Gene expression correlates of clinical prostate cancer behavior

Dinesh Singh,[1,5,12] Phillip G. Febbo,[1,4,8,11,12] Kenneth Ross,[8] Donald G. Jackson,[10] Judith Manola,[3] Christine Ladd,[8] Pablo Tamayo,[8] Andrew A. Renshaw,[6,14] Anthony V. D'Amico,[7] Jerome P. Richie,[5] Eric S. Lander,[8,9] Massimo Loda,[1,6] Philip W. Kantoff,[1,4] Todd R. Golub,[2,8,13] and William R. Sellers[1,4,11,13]

[1]Department of Adult Oncology
[2]Department of Pediatric Oncology
[3]Department of Biostatistical Sciences
Dana-Farber Cancer Institute
[4]Department of Internal Medicine
[5]Department of Surgery/Urology
[6]Department of Pathology
[7]Department of Radiation Oncology
Brigham and Women's Hospital
Harvard Medical School, Boston, Massachusetts 02115
[8]Whitehead Institute/Massachusetts Institute of Technology, Center for Genome Research, Cambridge, Massachusetts 02139
[9]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
[10]Pharmaceutical Research Institute, Bristol-Myers Squibb, Inc., Princeton, New Jersey 08543
[11]Correspondence: phil_febbo@dfci.harvard.edu (P.G.F.) and william_sellers@dfci.harvard.edu (W.R.S.)
[12]These authors contributed equally
[13]These authors codirected this work
[14]Present address: Baptist Hospital of Miami, Miami, Florida 33176

Summary

Prostate tumors are among the most heterogeneous of cancers, both histologically and clinically. Microarray expression analysis was used to determine whether global biological differences underlie common pathological features of prostate cancer and to identify genes that might anticipate the clinical behavior of this disease. While no expression correlates of age, serum prostate specific antigen (PSA), and measures of local invasion were found, a set of genes was identified that strongly correlated with the state of tumor differentiation as measured by Gleason score. Moreover, a model using gene expression data alone accurately predicted patient outcome following prostatectomy. These results support the notion that the clinical behavior of prostate cancer is linked to underlying gene expression differences that are detectable at the time of diagnosis.

Introduction

Prostate cancer is the most common nondermatological cancer in the United States with an estimated 198,100 new cases and 31,500 deaths in 2001 (Greenlee et al., 2000). The adoption of screening based upon the measurement of the serum prostate specific antigen (PSA) has led to the earlier detection of prostate cancer where most tumors now appear confined to the prostate gland at presentation (Han et al., 2001). Early diagnosis provides an opportunity for curative surgery. However, up to 30% of men undergoing radical prostatectomy will relapse, often as a result of micrometastatic disease present at the time of surgery (Roberts et al., 2001a, 2001b). The challenge is to identify those patients at risk for relapse and to better understand the molecular abnormalities that define tumors at risk for relapse.

Several clinical features of prostate cancer including tumor stage (Jewett, 1975), degree of tumor cell differentiation or Gleason score (GS) (Gleason, 1966), and the serum PSA (Stamey et al., 1987) are used in routine clinical practice to separate men into groups at low, intermediate, and high risk for tumor recurrence following local therapy. However, the majority of patients who now undergo prostatectomy have low to intermediate risk clinical features, and determining the prognosis for these patients remains difficult.

The utility of existing prognostic factors might be limited because they largely measure tumor differentiation and bulk but

SIGNIFICANCE

Improved patient stratification can allow the rational application of current treatments and the selected testing of novel therapeutics in patient populations most likely to benefit. Clinical features including Gleason score, tumor stage, and serum prostate specific antigen (PSA) are used to assess relapse-risk in men with prostate cancer. Such parameters are less useful in guiding therapy for men having intermediate risk disease, 30% of whom recur following local therapy. Our data suggest that expression-based models may help to identify patients at greatest risk for recurrence and thus facilitate the rational application of current therapies. Furthermore, the association of specific genes' expression, such as platelet-derived growth factor β (PDGFRβ), with outcome raises the possibility that expression analysis may prove useful in selecting patients for emerging mechanism-based therapeutics.

do not otherwise sample the underlying biological properties that likely drive tumor behavior. Attempts to explore genetic correlates of tumor behavior have found alterations in a number of candidate genes associated with prostate cancer progression, including loss of p53, amplification of myc, loss of p27, and loss of PTEN (reviewed in Sellers and Sawyers, 2001). However, no single gene has been shown to have sufficient prognostic utility to warrant clinical implementation.

Recently, genomic methodologies have been used to discover consistent gene expression patterns associated with a given histological or clinical phenotype (Golub et al., 1999; Perou et al., 2000; van't Veer et al., 2002). Here, gene expression patterns from 52 tumor and 50 normal prostate specimens were studied in order to ask whether such patterns could be used to predict common clinical and pathological phenotypes relevant to the treatment of men diagnosed with this disease. In addition to expression patterns that correlated with GS and with the distinction of tumor from normal, an expression-based model was built that accurately predicted patient outcome. These data suggest that it may be possible to predict the clinical behavior of prostate cancer based upon gene expression analysis of primary tumors. Such prediction strategies, if generalizable, would allow for the rational application of additional postsurgical therapeutics to high-risk individuals.

Results

Tumor versus normal classification

To investigate whether robust gene expression differences could be found that distinguished common clinical and pathological features of prostate cancer, 235 radical prostatectomy specimens were analyzed from patients undergoing surgery between 1995 and 1997. Of these samples, 65 had tumor on opposing sides of the tissue specimen. High-quality expression profiles were successfully derived from 52 of these prostate tumors and 50 nontumor prostate samples (referred to as normal hereafter) using oligonucleotide microarrays containing probes for approximately 12,600 genes and ESTs (raw data available at http://www-genome.wi.mit.edu/MPR/prostate). The clinical and pathological features of the 52 patients and tumors included in this study were indistinguishable from those of all patients treated with radical prostatectomy during the collection period (Table 1).

Genes were ranked according to their differential expression across the two classes (tumor versus normal) using a variation of a signal-to-noise metric (S2N) (Golub et al., 1999). The statistical significance of these gene expression correlations was determined by comparing the observed correlations to the results derived from 1000 permutations of the class labels (tumor or normal). This analysis indicated that 317 genes had higher expression in the tumor samples ($p \leq 0.001$) whereas 139 genes were more highly expressed in normal prostate samples ($p \leq 0.001$) (Supplemental Figure S1 [see Supplemental data, below]).

Gene expression differences between tumor and normal prostate samples have been previously reported (Chetcuti et al., 2001; Dhanasekaran et al., 2001; Luo et al., 2001; Welsh et al., 2001); however, the feasibility of using such differences to predict the identity of prostate samples has not been tested. To this end, we built predictors using a *k*-nearest neighbor (*k*-NN) supervised machine learning algorithm. Models that utilized 4 or more genes classified samples with greater than 90% accuracy in leave-one-out cross-validation testing ($p < 0.001$ as measured by permutation testing) (Suppl. Figure S2A and S2B). The 4-gene and 16-gene models were tested on an independent data set of 8 normal and 27 tumor prostate samples provided by G. Hampton (Welsh et al., 2001). Despite a nearly 10-fold difference in overall microarray intensity between these datasets (see Supplemental Experimental procedures [below]), the classifier performed with relatively high accuracy (4-gene model 77%; 16-gene mode 86%; $p < 0.05$, Fisher's exact test) (Suppl. Figure S3). Thus, expression differences can be used to predict the identity of unknown prostate samples and these gene expression differences are conserved across independent data sets.

Prediction of pathologic features of prostate cancer

In order to ask whether gene expression patterns exist that describe and or predict the differences in clinical behavior apparent among prostate tumors, the expression patterns within the 52 tumors were analyzed. Correlations between gene expression and known clinical and pathological parameters were determined for dichotomous variables (e.g., the presence or absence of capsular penetration, perineural invasion, or positive surgical margins), as well as for factors treated as continuous variables (e.g., patient age, serum PSA, and GS). Statistical significance was determined by comparing the observed correlations to those correlations measured in randomly permuted datasets. With the exception of GS (see below), no statistically significant gene expression correlates of these clinical and pathological features were observed (see Suppl. Figure S4). Specifically, no expression signature discriminated between locally invasive and noninvasive phenotypes (e.g., capsular penetration, positive surgical margins, and perineural invasion). Thus, while these features are often associated with different clinical outcomes, they are not reflected by global gene expression differences.

A gene expression signature of GS, however, was detectable. Fifteen genes had expression positively correlated with GS (Type I) and 14 genes had expression negatively correlated with GS (Type II) beyond what would be expected by chance alone ($p \leq 0.001$) (Figure 1 and Suppl. Figure S4). As these genes were the most positively and negatively correlated with GS, when used in hierarchical clustering, the 29 Type I and Type II were, as expected, separated into two groups (Figure 1). The correlation of these genes with GS and their coordinate expression in tumors, nonetheless, may have occurred by random chance alone in the initial dataset. However, when the same 29 genes were used to drive hierarchical clustering of the independent data set, Type I and Type II genes remained highly cosegregated suggesting that this coexpression is reproducible ($p < 0.0001$) (Suppl. Figure S5).

Strikingly, in both data sets, while most high-grade tumors expressed the Type I genes, a subset of intermediate grade tumors also expressed many of the Type I genes (Figure 1 and Suppl. Figure S5). This indicates that some tumors of intermediate histological grade share the gene expression signature of higher grade tumors. Thus, the coexpression of these genes may identify tumors that are of intermediate histological grade, yet share the molecular phenotype of high-grade tumors.

Prediction of clinical outcome

In this data set, 21 patients were evaluable with respect to recurrence following surgery with 8 patients having relapsed

Table 1. Clinical and pathological features

Variable		Study group	All	p	Recurrent	Nonrecurrent	p
#Patients		52	393		8	13	
Age	Median	58.5	61	0.32	58.5	60.0	0.74
	Range	47–72	40–79		51–72	47–72	
PSA	Median	6.3	6.7	0.62	6.8	6.3	0.64
	Range	1.0–27.8	0.7–46.0		5.0–24.3	3.6–18.0	
Gleason Score (Clinical)	2–6	19 (37%)	190 (51%)	0.10	2 (25%)	6 (46%)	0.45
	7	29 (56%)	146 (39%)		5 (63%)	6 (46%)	
	8–10	4 (8%)	34 (9%)		1 (12%)	1 (8%)	
	Unknown		23				
Gleason Score (Sample)	2–6	24 (46%)		0.46*	1 (12%)	7 (54%)	0.01
	7	22 (42%)			3 (38%)	6 (46%)	
	8–10	6 (12%)			4 (50%)	0	
Clinical Stage	T1–T2a	38 (88%)	285 (79%)	0.10	5 (100%)	10 (91%)	1.00
	T2b	5 (12%)	24 (7%)		0	9 (9%)	
	≥T2c	0	50 (14%)		0	0	
	Unknown	9	34		3	2	
Pathologic Stage	T2a	7 (13%)	49 (15%)	0.15	1 (13%)	2 (15%)	0.90
	T2b	25 (48%)	189 (58%)		4 (50%)	3 (38%)	
	T3a	16 (31%)	74 (23%)		2 (25%)	4 (31%)	
	T3b	4 (8%)	12 (4%)		1 (13%)	2 (15%)	
	T4a	0	2 (1%)		0	0	
	Unknown		53				
Gland Vol.	Median	51.75	53.0	0.89	67.5	50.0	0.15
	Range	35–191	18–191		35.5–191	35–169	
Ext. Cap.	No	32 (62%)	239 (73%)	0.10	5 (63%)	7 (54%)	1.00
	Yes	20 (38%)	88 (27%)		3 (37%)	6 (46%)	
	Unknown		66				
SVInv.	No	49 (94%)	315 (96%)	0.44	7 (88%)	12 (92%)	1.00
	Yes	3 (6%)	12 (4%)		1 (12%)	1 (8%)	
	Unknown		66				
Pos. Mar.	No	39 (75%)	283 (76%)	0.86	5 (62%)	7 (54%)	1.00
	Yes	13 (25%)	89 (24%)		3 (38%)	6 (46%)	
	Unknown		21				

PSA, serum prostate specific antigen; Vol., Volume; SV Inv., seminal vesicle invasion; Ext Cap., Extension through capsule; Pos. Margin, positive surgical resection margin. Gleason Score (Clinical) indicates the Gleason Score recorded from the radical prostatectomy specimen. Gleason Score (Sample) indicates the Gleason Score of the frozen sections from the tumor specimens used in RNA preparation. *p value resulting from the comparison of the Gleason Score (Sample) of the 52 tumors to the Gleason Score (Clinical) from the entire population.

(defined as two successive PSA values > 0.2 ng/ml) and 13 patients having remained relapse free for at least 4 years. While these two groups did not differ with respect to the Clinical GS, serum PSA, or tumor stage, the GS of the sections adjacent to tissue used for RNA extraction was ≥8 in a greater proportion of recurrent patients (4/8 versus 0/13) (Table 1).

While no single gene was statistically associated with recurrence (at $p = 0.05$) (data not shown), when a *k*-NN classification approach was applied, a 5-gene model with 2 nearest neighbors ($k = 2$) reached 90% accuracy in predicting recurrence during leave-one-out cross validation. When Kaplan-Meier survival analysis was performed based upon the predicted outcome, the results compared favorably with known prognostic indicators in this data set (Figure 2B). However, the standard Kaplan-Meier log-rank statistic, while demonstrating a difference in the survival curves, does not account for the multiple hypothesis testing that occurred during model optimization. To further assess the statistical significance of this prediction model, we performed 1000 permutations of the class labels (recurrence versus nonrecurrence), and for each permutation attempted to find multigene expression classifiers using the same range of gene numbers. Only 37 of the 1000 permutations yielded models whose accuracy matched or exceeded 90%. Thus, the likelihood of matching the success of the observed 5-gene model simply by chance alone was estimated at $p = 0.037$ (Figure 2A).

While there were too few tumor samples to allow for multivariate analysis, as mentioned above, only the Sample GS was significantly different between patients who recurred and those who did not recur (Figure 2B and Table 1). Nonetheless, 4 recurrent tumors were of intermediate grade (GS ≤ 7) raising the possibility that gene expression-based models might provide additional prognostic information not currently described by existing clinical and pathological parameters.

The genes that were used by the 5-gene outcome predictors during leave-one-out cross validation are shown in Figure 3. The top 5 genes were each used in over half of the models, and included chromogranin A, platelet-derived growth factor receptor β (PDGFRβ), HOXC6, inositol triphosphate receptor 3 (IPTR3) and sialyltransferase-1.

Discussion

There is an immediate need for robust prognostic markers capable of identifying patients at risk of relapse following local therapy; conventional and experimental therapeutics could then be focused on this subpopulation, rather than the general population of prostate cancer patients, 70% of whom are cured by surgery alone.

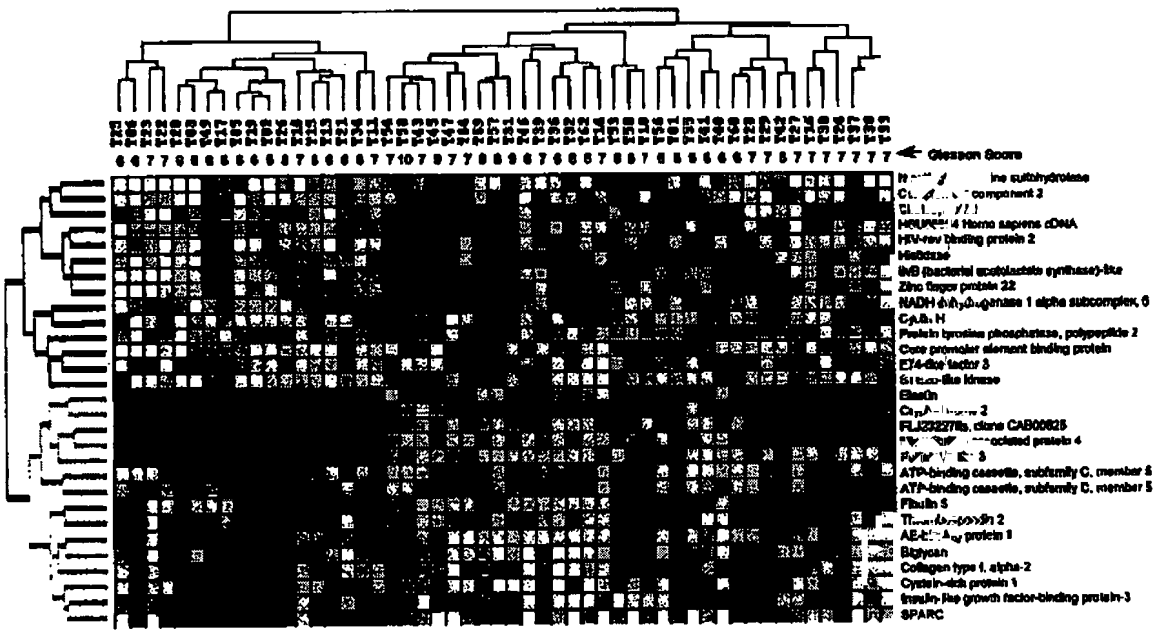

Figure 1. Gene expression correlates of Gleason score

Hierarchical clustering of tumors and the 29 genes statistically associated with GS. Genes and samples are shown as ordered by Gene Cluster and Treeview. The expression of each gene in each sample is represented by the number of standard deviations above (red) or below (blue) the mean for that gene across all 52 samples.

Our analysis revealed global gene expression differences that were sufficiently robust to distinguish tumor from normal in both training and validation sets. While the level of accuracy (86%–92%) is not sufficient to *replace* histological examination, these molecular markers may be useful adjuncts to morphology-based diagnostics. In addition, while certain genes differentially expressed between normal and tumor prostate specimens in microarray experiments have been correlated with outcome in large data sets (Dhanasekaran et al., 2001), in our data such differentially expressed genes were not highly correlated with outcome.

Among prognostic factors for prostate cancer, serum PSA and measures of local invasion were not associated with robust gene expression signatures. The lack of an associated gene expression signature does not exclude the possibility that such signatures exist. It is possible that analyzing genes beyond those present on the current microarray, or extending the experiment to larger data sets might reveal such patterns. However, there was a readily detectable and statistically significant signature of GS. The expression pattern of these genes separated tumors into distinct groups during hierarchical clustering in both our initial and in a validation data set and grouped some of the intermediate-grade tumors with high-grade tumors. In so doing, this set of genes may help identify a subset of histologically intermediate-grade tumors that may have more aggressive clinical behavior.

In this data set, GS was associated with patient outcome (Table 1 and Figure 2B); however, only two of the genes correlated with GS (IGFBP-3 and COL1A2) contributed to the outcome prediction model (Figure 3). Instead, genes whose expression was not correlated with GS were the most frequently used in the outcome prediction model, suggesting that while GS is associated with patient outcome, GS-independent markers and determinants of prostate cancer behavior exist.

Attempts to build a gene expression-based predictor of recurrence following prostatectomy led to a model that correctly predicted the outcome of 19 of the 21 evaluable patients in this study. While the result reached statistical significance based on permutation testing ($p = 0.037$), the performance of the model may be due, at least in part, to overoptimization. As such, this is a preliminary model and larger datasets will be required to reach model stability, to minimize the possibility of model overfitting and ultimately allow the independent validation of such a predictor.

Despite these limitations, the identity of the genes comprising the outcome prediction model support the existence of measurable outcome determinants for prostate cancer recurrence. For example, chromogranin A, one of the 5 genes most frequently used in the prediction model, has previously been associated with poor outcome in prostate cancer (Theodorescu et al., 1997). The utility of PDGFRβ expression in the recurrence predictor is also intriguing in light of the recent observations that PDGFR (α and β) are expressed in advanced prostate cancer (Chott et al., 1999). The successful prediction of patient outcome will ultimately lead to improved decision making regarding current therapeutic options and the rational selection of patients at high risk for relapse for clinical trials testing adjuvant therapeutics. Furthermore, the identification of genes whose expression drives outcome and whose protein products are tractable targets for small molecules may contribute to the development and selective application of novel mechanism-based treatments. Ongoing trials of Gleevec, an inhibitor of the abl,

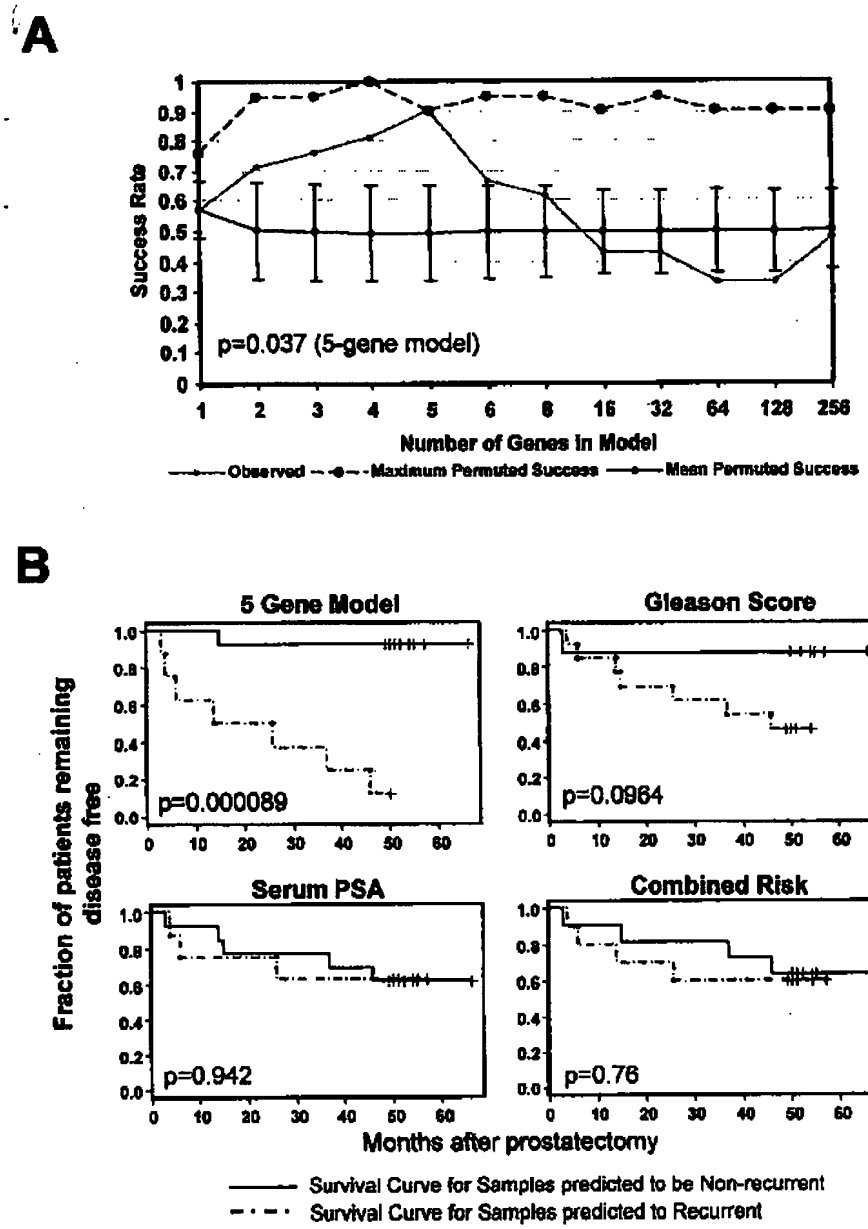

Figure 2. Outcome prediction models

A: The success rates of models predicting outcome. Leave-one-out cross validation was used to build outcome prediction models (recurrent versus nonrecurrent) using from 1 to 256 genes. The x axis indicates number of genes used in model building, and the y axis indicates the frequency of success. Shown is the number of correct predictions divided by the total number of predictions (red line) in the observed data using leave-one-out cross validation. The mean success rate ± the standard deviation (bottom dashed line) and maximum success rate (top dashed line) obtained using the permuted data is shown.

B: Disease-free survival of patients stratified based on the 5-gene model, GS, serum PSA, or combined risk. Kaplan-meir analysis was used to plot the fraction of at-risk patients remaining free of disease (y axis) at the indicated time after prostatectomy (x axis). Shown is patient stratification based on the 5-gene model, GS (≤ 6 versus ≥ 7), serum PSA (< 10 versus PSA ≥ 10), and a combination of GS, serum PSA, and surgical stage (low and intermediate versus high risk). High risk was defined as a GS > 7, PSA ≥ 20, and surgical stage T3 or higher. The remaining samples were considered low or intermediate risk. P values were calculated using a log-rank test (Mantel-Hwenszel test).

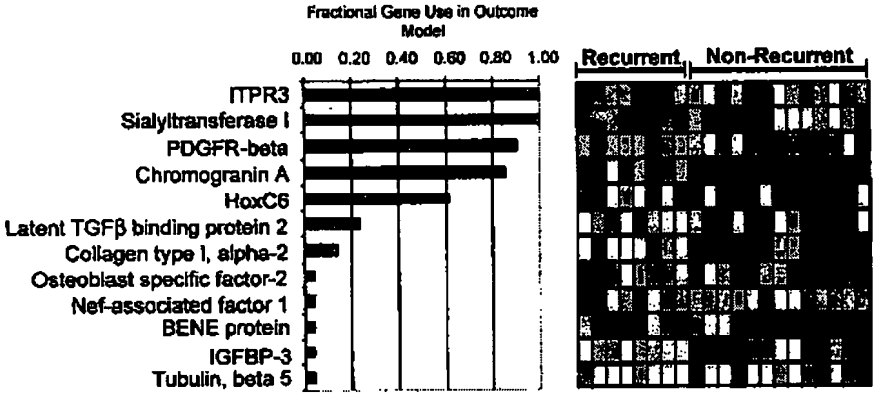

Figure 3. Genes used to build an outcome prediction model

The genes most commonly used in the 5-gene model are shown as described for Figure 2B. The expression of each gene (rows) in each recurrent or nonrecurrent sample (columns) is represented by the number of standard deviations above (red) or below (blue) the mean for that gene across all 21 samples.

kit, and PDGFR tyrosine kinases, in prostate cancer will test the hypothesis that PDGFRβ falls into this latter category.

The samples in this study were derived from patients diagnosed after the widespread adoption of PSA screening. As such the findings in this study are expected to be relevant to patients diagnosed today. Clearly, larger, confirmatory studies will be required prior to the implementation of any changes in the clinical care of patients with prostate cancer. Nevertheless, these studies provide evidence that the clinical phenotypes and behavior of prostate cancer can be anticipated by the analysis of the gene expression profiles.

Experimental procedures

Prostate tissue samples

From 1995 to 1997 samples of prostate tumors and adjacent prostate tissue not containing tumor (referred to as "normal") were collected from patients undergoing radical prostatectomy at the Brigham and Women's Hospital. From 235 "tumor" samples, 65 had cancer present on opposing sides of the OCT embedded specimens. Each of these samples was reviewed by a single pathologist to determine the "Sample" GS. Other pathological features from the radical prostatectomy specimens included in this analysis as well as from all contemporary prostatectomies were abstracted from the pathology reports including the "Clinical" GS.

The Wilcoxon rank sum test (Wilcoxin, 1945) and Fisher's exact test (Cox, 1970) was used to test for differences in continuous variables and dichotomous variables, respectively, between the study sample and all patients treated from 1993 to 1997 and between patients who recurred and those who did not. Tests for differences in these groups on ordered, categorical variables were done using the methods described by Mehta (Mehta and Patel, 1984). Kaplan-Meir survival plots and log-rank statistics (Mantel-Haenszel test) were generated using the S-Plus statistical software package (Insightful Corp).

Gene expression measurements

Total RNA extraction, generation of labeled cRNA, fragmentation, hybridization to U95Av2 arrays (Affymetrix) and wash steps were performed as previously described (Bhattacharjee et al., 2001; Golub et al., 1999). Raw expression values were normalized to the median array intensity and thresholds were set at 10 and 16,000 units. Genes whose expression varied less than 5-fold between any two samples in any given experiment were removed.

Gene ranking, class prediction by *k*-nearest neighbors, and permutation testing for dichotomous variables

Gene expression differences associated with a particular dichotomous class distinction were measured and ranked using a variation of the S2N statistic as previously described (Golub et al., 1999). Measured S2N values were compared to calculated S2N values obtained in 1000 data sets where a given class label was randomly permuted (Good, 2000). For these comparisons, p values represent the frequency at which the S2N statistic from randomly permuted data exceeded the measured S2N statistic.

K-nearest neighbor (*k*-NN) class prediction models were built as previously described (Pomeroy et al., 2002). Briefly, for each class distinction tested, after exclusion of one sample, genes were ranked using the S2N metric derived from the remaining samples. The Euclidean distances (ED) between the withheld sample and the remaining samples were calculated using a given number of genes (as in the figures). The identity of the left-out sample was predicted based upon the class membership of the *k*-closest samples weighted by the reciprocal of the EDs. P-values were then assigned based upon the frequency with which models generated and tested on 1000 randomly permuted data sets performed better than models generated using the observed data.

Correlation of gene expression with continuous variables

The Pearson coefficient was used to measure correlations between gene expression patterns and patient age, serum PSA, and GS (treated as a continuous variable). Pearson coefficients were also used to measure the same correlations in data sets where the sample label of each variable tested (age, PSA, or GS) was randomly permuted 10,000 times. P-values were then assigned based upon the frequency with which Pearson coefficients generated from randomly permuted data exceeded those generated from the observed data.

Independent prostate expression data used for validation

Oligonucleotide array-based expression data (Affymetrix Hum95Av2) and clinical data for 8 normal and 27 prostate tumors were provided G. Hampton as a validation set (Welsh et al., 2001). Global differences between the initial and validation data sets were quantified by determining the means of the mean array intensities. Validation of the tumor normal prediction models and of the coexpression observed for the genes highly correlated to GS was performed as described in the Supplemental experimental procedures (see Supplemental data, below).

Supplemental data

Supplemental experimental procedures and Figures S1–S5 can be found at http://www.cancercell.org/cgi/content/full/1/2/203/DC1 and at http://www-genome.wi.mit.edu/MPR/prostate.

Acknowledgments

The authors thank David Livingston and Matthew Meyerson for their assistance in this research and for critical reading of this manuscript, and Garret Hampton for providing the independent prostate cancer expression data set. This work was supported by the National Cancer Institute (CA84995) and in part by the Gelb Center at DFCI, Bristol-Meyers Squibb, Affymetrix and Millennium Pharmaceuticals. P.G.F. also receives support from the National Cancer Institute (CA89031) and from CaP CURE. W.R.S. also receives support from the Damon-Runyon Lilly Clinical Investigator Award and CaP CURE.

Received: January 10, 2002
Revised: March 1, 2002

References

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. USA *98*, 13790–13795.

Chetcuti, A., Margan, S., Mann, S., Russell, P., Handelsman, D., Rogers, J., and Dong, Q. (2001). Identification of differentially expressed genes in organ-confined prostate cancer by gene expression array. Prostate *47*, 132–140.

Chott, A., Sun, Z., Morganstern, D., Pan, J., Li, T., Susani, M., Mosberger, I., Upton, M.P., Bubley, G.J., and Balk, S.P. (1999). Tyrosine kinases expressed in vivo by human prostate cancer bone marrow metastases and loss of the type 1 insulin-like growth factor receptor. Am. J. Pathol. *155*, 1271–1279.

Cox, D. (1970). Analysis of Binary Data (London: Machuen and Co.).

Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A., and Chinnaiyan, A.M. (2001). Delineation of prognostic biomarkers in prostate cancer. Nature *412*, 822–826.

Gleason, D. (1966). Classification of prostatic carcinomas. Cancer Chemother. Rep. *50*, 125–128.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531–537.

Good, P. (2000). Permutation tests: A practical guide to resampling methods for testing hypotheses, Second Edition, S.S.I. Statistics (New York: Springer-Verlag).

Greenlee, R.T., Murray, T., Bolden, S., and Wingo, P.A. (2000). Cancer statistics, 2000. CA Cancer J. Clin. *50*, 7–33.

Han, M., Partin, A.W., Piantadosi, S., Epstein, J.I., and Walsh, P.C. (2001). Era

specific biochemical recurrence-free survival following radical prostatectomy for clinically localized prostate cancer. J. Urol. *166*, 416–419.

Jewett, H.J. (1975). The present status of radical prostatectomy for stages A and B prostatic cancer. Urol. Clin. North Am. *2*, 105.

Luo, J.H., Yu, Y.P., Cieply, K., Lin, F., Deflavia, P., Dhir, R., Finkelstein, S., Michalopoulos, G., and Becich, M. (2001). Gene expression analysis of prostate cancers. Mol. Carcinog. *33*, 25–35.

Mehta, C., and Patel, A.T. (1984). Exact significance testing for ordered categorical data. Biometrics *30*, 819–825.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Pomeroy, S.L., Tamayo, P., Gaasenbeek, M., Sturla, L.M., Angelo, M., McLaughlin, M.E., Kim, J.Y., Goumnerova, L.C., Black, P.M., Lau, C., et al. (2002). Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature *415*, 436–442.

Roberts, S.G., Blute, M.L., Bergstralh, E.J., Slezak, J.M., and Zincke, H. (2001a). PSA doubling time as a predictor of clinical progression after biochemical failure following radical prostatectomy for prostate cancer. Mayo Clin. Proc. *76*, 576–581.

Roberts, W.W., Bergstralh, E.J., Blute, M.L., Slezak, J.M., Carducci, M., Han, M., Epstein, J.I., Eisenberger, M.A., Walsh, P.C., and Partin, A.W. (2001b). Contemporary identification of patients at high risk of early prostate cancer recurrence after radical retropubic prostatectomy. Urology *57*, 1033–1037.

Sellers, W., and Sawyers, C. (2001). Molecular abnormalities in prostate cancer. In Prostate Cancer: Principles & Practice, P. Kantoff, P. Carroll, and A. D'Amico, eds. (Philadelphia: Lippincott Williams & Wilkins), pp. 53–76.

Stamey, T.A., Yang, N., Hay, A.R., McNeal, J.E., Freiha, F.S., and Redwine, E.A. (1987). Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N. Engl. J. Med. *317*, 909.

Theodorescu, D., Broder, S.R., Boyd, J.C., Mills, S.E., and Frierson, H.F., Jr. (1997). Cathepsin D and chromogranin A as predictors of long term disease specific survival after radical prostatectomy for localized carcinoma of the prostate. Cancer *80*, 2109–2119.

van't Veer, L., Dai, H., Vijver, M.v.D., He, Y., Hart, A., Mao, M., Peterse, H., Kooy, K.v.D., Marton, M., Witteveen, A., Schreiber, G., Kerkhoven, R., Roberts, C., Linsley, P., Bernards, R., and Friend, S. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Welsh, J.B., Sapinoso, L.M., Su, A.I., Kern, S.G., Wang-Rodriguez, J., Moskaluk, C.A., Frierson, H.F., Jr., and Hampton, G.M. (2001). Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res. *61*, 5974–5978.

Wilcoxin, F. (1945). Individual comparisons by ranking methods. Biometrics *1*, 80–83.